Guest Editors’ Introduction: Intelligence in the Cloud

Song Guo, The Hong Kong Polytechnic University
Victor Leung, The University of British Columbia
Xin Yao, Southern University of Science and Technology
rtificial intelligence (AI), since its birth in the 1950s, has been believed to be the key to our civilization’s brightest future. To pursue the vision of AI, various machine learning approaches, e.g., deep learning, supervised learning, unsupervised learning, reinforcement learning, etc., have been proposed. The coming big data era has enthusiastically renewed the call and focus for advanced machine learning technologies to extract knowledge from large data pools. With its rich resource provisioning, cloud computing is widely regarded as an ideal platform to facilitate resource-intensive machine learning so as to enable intelligence in the cloud. Integrating intelligence into the cloud is without doubt a promising development trend to both cloud computing and AI.

In hardware support for providing intelligence in the cloud, many companies have been designing specialized chips for AI, especially for neural networks, because powerful computation is the key of AI. These chips are widely used in every corner of clouds from cores to edges. GPUs and tensor processing units (TPUs) are the two most powerful AI chips. In 2017, NVIDIA released the Tesla V100 GPU with new Volta architecture, which particularly incorporates Tensor Cores into streaming multiprocessors as well as other general improvements. In the same year, Google announced TPU 2.0 (Cloud TPU) aiming to connect the tensor-specific computation into larger systems, e.g., the Google Compute Engine. For mobile-edge devices, Apple released the Apple Neural Engine, a module of System-on-Chip (SoC), to process AI tasks. Qualcomm and Huawei also designed their own AI modules in SoC. In software support, many deep learning frameworks have been deployed to the cloud, such as Tensorflow (Google), Caffe2 (Facebook), CNTK (Microsoft), MXNet (Amazon), Deeplearning4j, etc. Deeplearning4j can be integrated with Hadoop and Spark and is designed to be used in business environments on distributed GPUs and CPUs.

On the other hand, AI techniques have also been widely used in resource management in the cloud. For example, reinforcement learning is used to improve job scheduling for spark streaming. Knowledge-Defined Networking is proposed as a new paradigm that accommodates and exploits Software-Defined Networking, Network Analytics, and AI for datacenter networking by extracting knowledge from network logs.

This special issue is intended to introduce the state-of-the-art, open research challenges, new solutions, and applications for intelligence in the cloud computing. Specifically, we choose three high-quality papers for this special issue, covering different aspects of resource management, machine learning based framework, and a Blockchain-based mechanism that enables intelligent cloud computing.

We are still at the early stage of integrating intelligence into cloud computing. The selected articles in this special issue show us a snapshot of recent developments in this area and help our readers identify other challenges to be addressed by both the research community and industry. Finally, we would like to acknowledge the great support from Mazin Yousif, the current Editor-in-Chief of IEEE Cloud Computing magazine, Beverly Lindeen, Managing Editor of IEEE Cloud Computing magazine, and other IEEE Computer Society publication staff.

SONG GUO is a full professor at the Department of Computing, The Hong Kong Polytechnic University. His research interests are mainly in the areas of big data, cloud computing, green communication and computing, and distributed systems. His work was included in 21st Annual Best of Computing—Notable Books and Articles in Computing of 2016 by ACM Computing Reviews. He also received 5 best paper awards from IEEE/ACM conferences and the IEEE Systems Journal Annual Best Paper Award of 2017. Dr. Guo currently serves on editorial boards of several prestigious journals, including IEEE Transactions on Emerging Topics in Computing, IEEE Transactions on Sustainable Computing, IEEE Transactions on Green Communications and Networking, and IEEE Communications. He is an active volunteer as general/TPC chair for more than 20 international conferences, chair/vice-chair for several IEEE technical committees and SIGs, and a keynote speaker and panelist for many domestic and international conferences. He is a senior member of the IEEE, a senior
VICTOR LEUNG is currently a professor and holds the position of TELUS Mobility Research Chair in Advanced Telecommunications Engineering with the Department of Electrical and Computer Engineering, The University of British Columbia. He has been involved in telecommunications research with a focus on wireless networks and mobile systems for more than 30 years, which has resulted in more than 1000 journal and conference papers coauthored with his students and collaborators, including several papers that have won best paper awards. Dr. Leung has contributed to the organization and technical program committees of numerous conferences. He was a distinguished lecturer of the IEEE Communications Society. He has contributed to the editorial boards of many journals, including the IEEE Transactions on Computers, the IEEE Transactions on Wireless Communications, the IEEE Transactions on Vehicular Technology, IEEE Wireless Communications Letters, and the IEEE Journal on Selected Area in Communications. He received an APEBC Gold Medal in 1977 as the head of the graduating class of the Faculty of Applied Science, a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship from 1977–1981, an IEEE Vancouver Section Centennial award in 2011, and a UBC Killam Research prize in 2011. He is a fellow of The Royal Society of Canada, the Engineering Institute of Canada, and the Canadian Academy of Engineering. He is a registered professional engineer in the Province of British Columbia. Contact him at vleung@ece.ubc.ca.

XIN YAO is a chair professor of Computer Science at the Southern University of Science and Technology (SUSTech) in Shenzhen, China, and a part-time professor at the University of Birmingham, UK. He is an IEEE Fellow and a distinguished lecturer of the IEEE Computational Intelligence Society (CIS). He previously served as the Editor-in-Chief (2003–08) of the IEEE Transactions on Evolutionary Computation and the president (2014–15) of IEEE CIS. His main research interests include evolutionary computation and ensemble learning, especially online ensemble learning and class imbalance learning, and their applications in software engineering and fault diagnosis. His papers won the 2001 IEEE Donald G. Fink Prize paper award, 2010, 2016, and 2017 IEEE Transactions on Evolutionary Computation outstanding paper awards, 2010 BT Gordon Radley Award for best author of innovation (Finalist), 2011 IEEE Transactions on Neural Networks Outstanding Paper award, and many other best paper awards. He won the prestigious Royal Society Wolfson Research Merit award in 2012 and the IEEE CIS Evolutionary Computation Pioneer award in 2013. Contact him at X.Yao@cs.bham.ac.uk.